

Automated Document Conversion

Master Plan

April 1995

Version 1.0

Office of the Assistant Secretary of Defense (Command, Control, Communications, and Intelligence/Information Management)

(Cleared for open publication April 12, 1995 by OSD (PA))

EXECUTIVE SUMMARY

INTRODUCTION

This Master Plan provides strategic guidance for all automated document conversion (ADC) acquisitions within the Department of Defense (DoD). For the purposes of this Master Plan, ADC guidance focuses on conversion from paper/microform (analog) to digital (used by a computer) formats, although the need for digital-to-digital conversion is addressed. This Master Plan views document conversion as an activity within the records management business process.

The Master Plan follows three basic themes:

- follow existing policy
- use a commonly shared DoD business process that also provides practical business rules
- implement a flexible document conversion strategy based on technical standards that provides the Department with sharable converted documents

DoD CONVERSION ENVIRONMENT

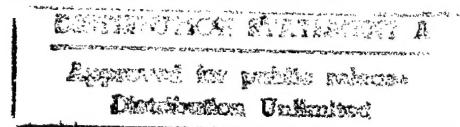
Establishing an effective digital document environment is a complex undertaking. Implementation of a digital document environment requires everyone involved to understand the business processes and workflow, to accept the process and technology, and to practice sound records and document management. Management must support the installation of a well-designed technology infrastructure and a financial plan to maintain the new digital environment over the long term.

Deciding what documents to convert is crucial to any conversion project. Military mission and business requirements, and a business case that clearly articulates the functional and economic benefits anticipated from conversion, will guide conversion decisions. For example, conversion should be undertaken only if it improves the accessibility of organization records and reduces costs to store and preserve them.

Records management may be viewed as the foundation for automated document conversion. This Master Plan places emphasis on using existing DoD policy-it applies existing records management policy and life-cycle management (LCM) policy for automated information systems.

The approval policy for ADC acquisitions is based upon compliance with these policies and the specific standards and technical guidance addressed in this Plan. The DoD Components are responsible for certifying the need for current and planned ADC acquisitions.

CONVERSION STRATEGY



19970929 096

PM

New Text Document.txt

24 Sep 97

This paper was downloaded from the Internet.

Distribution Statement A: Approved for public release;
distribution is unlimited.

POC: Office of the Assistant Secretary
of Defense (C3I/Information Mgt)

DTC QUALITY INSPECTED 4

The converted document must meet specifications compatible with the functional application for which it was generated, but not be locked into a proprietary digital format. Given the size and duration of many conversion efforts, DoD is adopting a two-stage process. The first stage establishes an interoperable environment where original documents are converted to a non-proprietary and interoperable format, allowing information to be shared. The second stage recognizes the need for additional intelligence in the document and further processes the digitized document into more complex digital formats if required by the application.

From a technology and acquisition point of view, the Conversion Strategy seeks to:

- maximize the flexibility and utility of converted documents to DoD document users through the requirement to use *de jure* (e.g., ISO, ANSI, IEEE) technical standards
- ensure that the output of a conversion effort supports a DoD mission or business process in a cost-effective manner
- avoid unnecessary duplication in the acquisition of ADC hardware and/or software

The Department of Defense is committed to the establishment of a standards-based framework for defining technical architectures that will help ensure interoperability, portability, and scalability of its information systems. The standards identified as most suitable for DoD ADC systems will move the Department toward open/non-proprietary systems. All ADC systems must use these standards or provide a business case analysis for using proprietary standards as well as a migration strategy to open systems.

ROLES AND RESPONSIBILITIES

This Master Plan assigns roles and responsibilities for DoD officials involved with ADC. The assignments support the management philosophy of centralized DoD policy direction for ADC within ASD(C3I) and decentralized execution and implementation of ADC by the functional DoD Component officials.

CONVERSION GUIDES FOR PROGRAM MANAGERS

A series of guides will be issued to provide managers with an understanding of the technical issues that must be addressed through a document conversion life-cycle. Because immediate conversion priorities are for paper/microform conversion to digital documents, the guides will focus on this type of conversion.

TABLE OF CONTENTS

- 1. INTRODUCTION 1
- 1.1 BACKGROUND 1
- 1.1.1 Congressional Guidance 2
- 1.1.2 DoD Working Group 3
- 1.2 SCOPE 3
- 1.3 CONTEXT 3
- 1.4 RELATIONSHIP TO AUTOMATED DOCUMENT CONVERSION SYSTEM TEST 4
- 1.5 ORGANIZATION OF THE MASTER PLAN 5

- 2. DOD CONVERSION ENVIRONMENT 7
- 2.1 OVERVIEW 7
- 2.2 MOTIVATION TO DIGITIZE DOCUMENTS 7
- 2.2.1 Goals 9
- 2.2.2 ADC Costs (Short-term and Long-term) 9

- 2.2.3 DoD Business Requirements 10
 - 2.3 RECORDS MANAGEMENT FOUNDATION 10
 - 2.3.1 Managing Converted Documents 11
 - 2.3.2 Document Types 11
 - 2.3.3 Records Management Requirements for Converted Documents as Organization Records 12
 - 2.3.4 Records Management Requirements for Original Documents as Organization Records 12
 - 2.4 BUSINESS NEED JUSTIFICATION 13
 - 2.4.1 Approval of ADC Acquisitions 13
-

- 3. CONVERSION STRATEGY 15
 - 3.1 OVERVIEW 15
 - 3.2 GOALS 15
 - 3.3 THE CONVERTED DOCUMENT: DEFINITION 16
 - 3.3.1 The 'Convert Document' Activity 16
 - 3.3.2 Two-Stage Document Conversion Process 16
 - 3.4 STANDARDS GUIDANCE FOR CREATING THE CONVERTED DOCUMENT 18
 - 3.4.1 TAFIM Policy 18
 - 3.4.2 Standards for DoD Automated Document Conversion 19
 - 3.4.3 Marketplace (De Facto) Standards 21
 - 3.5 MANAGEMENT COMPONENT OF CONVERSION STRATEGY 22
 - 3.5.1 Requirements Determination 23
 - 3.5.2 Cost Justification 25
 - 3.5.3 Technical Capability 25
 - 3.5.4 Business Case Decision Table 25
-

- 4. ROLES AND RESPONSIBILITIES 29
 - 4.1 Assistant Secretary of Defense (C3I) 29
 - 4.2 Functional Officials 29
 - 4.3 DoD Components 29
 - 4.4 Commanders, Managers, Records Officers, and Document Custodians 30
-

- APPENDIX A - EXAMPLES OF BUSINESS REQUIREMENTS FOR AUTOMATED DOCUMENT CONVERSION A-1
 - APPENDIX B - CONVERT DOCUMENT PROCESS MODEL B-1
 - APPENDIX C - DEFINITIONS C-1
 - APPENDIX D - BASIC ELEMENTS OF ELECTRONIC DOCUMENT CONVERSION D-1
 - APPENDIX E - CONVERSION GUIDES FOR PROGRAM MANAGERS E-1
 - APPENDIX F - ACRONYMS F-1APPENDIX G - REFERENCES G-1
-

FIGURES

- FIGURE 1 - CONTEXT 4
- FIGURE 2 - INDUSTRY USE OF DOCUMENTS IN DIGITAL FORMS 8
- FIGURE 3 - REASONS FOR AUTOMATED DOCUMENT CONVERSION 8
- FIGURE 4 - INDUSTRY CONVERSION COST TRENDS 10
- FIGURE B-1 - CONVERSION CONTEXT DIAGRAM B-2
- FIGURE B-2 - CONVERSION PROCESS B-3
- FIGURE D-1 - THE BASIC IMAGING PROCESS D-1
- FIGURE D-2 - ASCII TEXT VS. IMAGING D-3

- FIGURE D-3 - DOCUMENT CAPTURE WORKFLOW D-6
- FIGURE D-4 - IMAGE STORAGE WORKFLOW D-7

TABLES

- Table I - Summary of ADC Standards 22
 - Table II - Business Case Decision Table 26
-

INTRODUCTION

This Master Plan sets forth the Department of Defense (DoD) management strategy for automated document conversion. The strategy centers on the effective conversion of documents to an electronic or digital format (used in a computer) and the management of converted documents throughout the life of the document. The Master Plan addresses information management, records management, business process improvement, and the management of document conversion hardware and software acquisition. The intent of this Master Plan is to provide consistent planning guidance for moving away from paper and microform documents to digitally based formats and technical standards. The following vision statement served as the guide for developing the conversion strategy.

DoD Automated Document Conversion efforts will strive to convert documents to standards-based formats that allow users, with differing functional requirements, to share the contents of documents and perform conversion based on a cost-justified business case, applying Corporate Information Management (CIM) principles and DoD policies and procedures.

The Master Plan follows three basic themes:

- follow existing policy
- use a commonly shared DoD business process that also provides practical business rules
- implement a flexible document conversion strategy based on technical standards that provides the Department with sharable converted documents

The Office of the ASD(C3I) will develop an action plan for follow-on activities (such as ADC management coordination and technical standards development) required to provide additional implementation guidelines. This Master Plan contains no milestones. Responsible functional and DoD Component officials, in accordance with DoD information management policy, define the military and business requirements, the timing, and the justification for automated document conversion.

BACKGROUND

Document conversion is typically viewed as converting paper documentary material to a digital form, although conversion from paper to microform, digital to microform, digital to digital, and other types of conversion are also common. Many managers in the business and technology communities perceive the need to replace analog (paper or microform) documents with digital documents.

The many technology-related marketing claims may have generated an artificial demand for conversion of DoD paper records to digital media and formats. Indeed, the perception of DoD "chasing" conversion technologies (e.g., imaging and electronic document management) may have helped to motivate the need for this Master Plan. Because of these rapidly changing and emerging technologies, the Department needs to exercise greater discipline in its acquisition of automated document conversion equipment, systems, and services. The acquisition policy is already in place.

The solution rests in all levels of DoD management with acquisition responsibility directing additional energies to answering questions related to document conversion business decisions. Those questions include the following:

- What documents, if any, must and will be converted? For what mission or business reason(s)?
- Who decides?
- How will those selected documents be converted?
- When can selected documents be converted? What criteria are needed to guide scheduling, staging, and phasing of document conversion? What workflow issues must be resolved during lengthy document conversion efforts?
- What are the short-term and long-term (measured by the lifetime of a document) costs of conversion? What benefits are qualified and quantified?

This Master Plan will answer some of these questions, and raise many more. Stimulating the reader to think about the topics and issues raised is a positive by-product of the Master Plan.

Mission and business environments using digitized documents (text, images, engineering drawings) can achieve improved administrative efficiency. Efficiency results from technology support for locating, retrieving, and disseminating digitized documents linked to workflow processes. However, the availability of digitized documents alone does not guarantee improved administrative efficiency. To be more efficient, re-engineering of business methods and processes should drive the digitization of information.

Congressional Guidance

Congress required, in Section 8114 of the Defense Appropriations Act of 1995, the Assistant Secretary of Defense for Command, Control, Communications, and Intelligence (ASD[C3I]) "to establish and implement a master plan for all acquisitions of automated document conversion systems, equipment, and technologies." Congress directed that no funds may be used "to develop technologies or to acquire new automated document conversion equipment, services, or systems which cost more than \$5,000,000 after April 1, 1995, unless such acquisitions are approved in advance by the Assistant Secretary or his designee." The Congressional conferees required that the DoD submit the Master Plan by April 1, 1995.

DoD Working Group

In October 1994, the ASD(C3I) convened an Automated Document Conversion Working Group. The purpose of the Working Group was to develop a Master Plan to address all future automated document conversion (ADC) systems, equipment, and technologies in accordance with Congressional guidance. The Working Group consisted of representatives from elements of the Office of the Secretary of Defense and the Military Departments and Defense Agencies.

SCOPE

This Master Plan provides:

- the information management business context within which documents (organization records) undergo conversion
 - a flexible document conversion strategy
 - policy guidance for acquisition management of ADC equipment, systems, and technologies.
-

CONTEXT

The information management business context for automated document conversion is illustrated in Figure 1. The Information Management functional area includes records management as a major

business process. Records management includes the creation, maintenance, transfer, and destruction of organization records.

Records management defines the policy and provides the structure for managing the life-cycle of information contained in records, regardless of medium of storage (i.e., paper, film, magnetic media). Electronic document management defines a subset of information handling concerned with capturing, retrieving, converting, storing, and disseminating digital forms of information. ADC management focuses on the conversion of documents (analog or digital) to the desired digital formats.

The Master Plan describes a document conversion strategy that is responsive to mission and business requirements. It also confirms the technical standards to which conversion output products (converted documents) must comply. Technical standards specifying document output formats will evolve. Conformance to the technical standards increases the utility of converted documents to the organization.

The Master Plan reiterates DoD automated information system (AIS) life-cycle management (LCM) policy and interprets that guidance for the management of all ADC system and service acquisitions within the Department. This ADC guidance focuses on conversion from paper/microform (analog) to digital formats, although the principles and guidelines apply equally to digital-to-digital and other types of document conversion.

FIGURE 1 - CONTEXT

RELATIONSHIP TO AUTOMATED DOCUMENT CONVERSION SYSTEM TEST

In Fiscal Year 1994, Congress directed the DoD to acquire and test an ADC system to determine if off-the-shelf conversion technologies could provide economical assistance to convert original technical documents to formats compatible with advanced computer applications. Two sets of conversion software were tested: one for engineering drawings and one for textual documents. Congress also directed the incorporation of "lessons learned" from the test into a broader policy document, the Automated Document Conversion Master Plan. The lessons apply to all conversion projects within the Department, including financial, personnel, and medical documents, as well as technical documentation.

Several lessons learned from the Automated Document Conversion System (ADCS) Test confirm the need for a consistent approach to document conversion. Specifically, the ADCS Test determined that:

- demand for conversion to intelligent forms exists
- active intelligent documents are required to support an integrated digital environment for DoD acquisition, procurement, and maintenance
- files and original documents presented for conversion vary widely in quality and complexity
- storage, indexing, and identifying intelligent documents in existing repositories could enhance the repositories' value beyond their original intended purposes by making those documents available throughout the DoD, thereby fostering increased access to important digital information.

ORGANIZATION OF THE MASTER PLAN

This Master Plan consists of four sections and a set of appendices that provide additional information pertinent to the use of the Master Plan.

- Section 1 (Introduction) describes the background, scope, and context of the Automated Document Conversion Master Plan.
- Section 2 (DoD Conversion Environment) provides management perspectives of automated document conversion. This section summarizes the motivation (mission and business needs) for document conversion, describes the foundation process and requirements for a document conversion activity, and interprets DoD records management and automated information system life-cycle management policies as they relate to automated document conversion.
- Section 3 (Conversion Strategy) provides a technology and acquisition perspective of automated document conversion. This section describes the DoD strategy for achieving a consistent approach

to automated document conversion, cites technical standards approved for DoD converted document formats, and addresses management and acquisition issues that may affect implementation of the strategy.

- Section 4 (Roles and Responsibilities) describes organizational roles and responsibilities applicable to automated document conversion.

DOD CONVERSION ENVIRONMENT

OVERVIEW

This section provides three management perspectives of automated document conversion:

- **Motivation.** Why should the DoD convert documents for use across the Department? What are the goals and costs?
- **Records Management Foundation.** ADC is viewed as a component of the records management business process. What are the accompanying laws, regulations, and DoD policy that apply?
- **Business Need Justification.** Using existing DoD policy, how will the Department manage and oversee ADC system and service acquisitions?

MOTIVATION TO DIGITIZE DOCUMENTS

Establishing an effective digital document environment is a complex undertaking. Implementation of a digital document environment requires everyone involved to understand the business processes and workflow, to accept the process and technology, and to practice sound records and document management. Management must support the installation of a well-designed technology infrastructure and a financial plan to maintain the new digital environment over the long term.

Organizations are vigorously pursuing the benefits of document conversion and imaging technology. Figure 2, based on an industry survey by the Gartner Group, illustrates the change in pace at which organizations are moving to digital document formats.

Deciding what documents to convert is crucial to any conversion project. Military mission and business requirements, and a business case that clearly articulates the functional and economic benefits anticipated from conversion, will guide conversion decisions. A recent survey of industry by the Gartner Group identified reasons for adopting imaging systems to support business cases (see Figure 3). (The terms "imaging system" and "ADC system" are used synonymously in this Master Plan.)

FIGURE 2 - INDUSTRY USE OF DOCUMENTS IN DIGITAL FORMS

FIGURE 3 - REASONS FOR AUTOMATED DOCUMENT CONVERSION

Goals

The following goals should be fulfilled when moving to a digital document environment:

- improve the accessibility of organization records to the DoD work force and the public
- reduce costs to store and preserve organization documents and improve service to document users
- ensure that the output of a conversion effort supports a DoD mission or business process in a cost-effective manner
- avoid unnecessary duplication in the acquisition of ADC hardware and/or software
- avoid repetitive or multiple conversions of the same document
- ADC Costs (Short-term and Long-term)

ADC costs have short-term and long-term components. The obvious short-term components include the cost of systems or services to perform document conversion, labor costs to determine the eligibility for conversion and perform quality control, and the cost to provide access mechanisms (e.g., workstations, servers, telecommunications networks). Since document conversion is not a one-time activity, long-term

cost components are less obvious. Long-term costs may be generated by the following limitations:

- The life of the document is longer than the medium on which it is stored. The document must be recopied to a newer storage medium.
- The life of the document is longer than the devices that store and read the medium (e.g., 20-year record stored on an IBM MagCard II device). The document must be read, then restored or reconverted to a new medium before the end of the access device's reparable life.

Figure 4 illustrates the order of magnitude cost differences to convert a one-page document using three different technical standard formats. It also illustrates the importance of selecting an output product format that satisfies the minimum business requirements. Based on the accelerated rate of technological change and the increasing adoption of more complex digital formats, costs can be expected to drop quickly in the near future. However, while Standard Generalized Markup Language (SGML) is a more expensive process than the other two, managers should not base their decision on the least expensive method, but on the one giving them the most flexibility for document management.

FIGURE 4 - INDUSTRY CONVERSION COST TRENDS

DoD Business Requirements

DoD holds vast amounts of organization records in paper, microform and digital formats. The majority of the holdings are on paper. No credible estimate of organization records eligible for document conversion exists. However, several conversion efforts are under way within DoD. Appendix A summarizes private-sector and approved DoD business requirements.

RECORDS MANAGEMENT FOUNDATION

The Department places special emphasis on using existing DoD policy. This section addresses ADC as it applies to existing records management policies for automated information systems.

The management of organization documents as records is outlined in 44 United States Code (USC) and 36 Code of Federal Regulations (CFR). Organization documents are records managed within approved records management procedures.

Records management is the information management business process within which document conversion is an element of maintaining records. Conversion does not alter access requirements or restrictions to documents under the Freedom of Information Act, Privacy Act, Information Security programs, or other forms of disclosure required by law. Conversion of documents requires responsible management of both the original and the converted documents.

Some documents such as medical, health, or life insurance records may be neither suitable nor eligible for conversion. The requirement for an original signature or witnessed authentication may determine suitability or eligibility for conversion.

Managing Converted Documents

A record consists of information, regardless of medium or structure, detailing the transaction of organization business. ADC is the migration of organization documents from one medium to another. Changing the medium does not change the document's status as a record. Both the original document and the converted document are managed in accordance with approved records management processes and procedures. ADC provides an excellent opportunity to manage organization records using an electronic records management system.

When records are converted to an electronic (digital) medium, that medium becomes an electronic records management system. That records system must be capable of:

- printing, presenting, and storing records in their native (original) format

- importing and exporting records between the system storage environment and paper, magnetic disk, optical disk, or other media that requirements justify
- converting system-stored records into standard (e.g., ASCII, TIFF, PDF, or SGML) formats
- storing multimedia information as a single record
- verifying the quality of records transferred
- implementing standardized approaches including cataloging and indexing techniques to facilitate expedient record retrieval and accessibility.

Document Types

Documents include but are not limited to the following physical types:

- forms
- handwritten documents
- type-written documents
- diagrams
- machine-created documents
- x-rays
- maps
- engineering drawings
- photographs

DoD uses a wide variety of document types. Several types of documents may be present in a functional document conversion requirement. Three examples follow:

- **Technical and Engineering.** These documents may include textual materials, geometric drawings, diagrams, photographs, and work cards. Conversion of technical documents may require large-sized equipment for paper-based engineering drawings and technical manual diagrams, microfilm scanners for engineering drawings preserved on aperture cards, and specialized scanners for x-ray films associated with ammunition management and welding operations.
- **Personnel and Finance.** Most personnel and finance documents consist of text and tabular materials, but may also include photographs and microform materials.
- **Medical.** These documents include text and tabular materials, charts, x-rays, and other types of recorded medical information.

Records Management Requirements for Converted Documents as Organization Records

When a converted document becomes the organization record, the converted document shall:

- be categorized and managed in accordance with approved records schedules
- not be maintained beyond the approved disposition
- meet requirements for management and transfer established by the National Archives and Records Administration

Documents will be indexed at the time of conversion to facilitate easy access and appropriate disposition. At least one index based on disposal date must be established. The index aids the identification and disposition of the converted documents in a timely and accurate manner.

Records Management Requirements for Original Documents as Organization Records

When the original document remains as the organization record, the converted documents shall be:

- managed to facilitate economical use of the information contained in the converted document
- kept only as long as economically necessary
- identified in the organization records schedule as either a working copy or other appropriate label that explicitly indicates the converted document is not the organization record

When the original document remains as the organization record, the original document shall be:

- managed in accordance with approved records schedules
- identified in the organization records schedule as the organization record copy

BUSINESS NEED JUSTIFICATION

Before acquiring an ADC system or service, a mission or business need must be justified using existing CIM principles and automated information system life-cycle management policy.

Approval of ADC Acquisitions

Section 8114 of the Department of Defense Appropriation Act, 1995, requires that, effective April 1, 1995, no funds shall be used to develop technologies or to acquire new automated document conversion (ADC) equipment, services or systems that cost more than \$5,000,000 unless such acquisitions are approved in advance by the ASD(C3I) or his designee.

The ASD(C3I) will issue procedures regarding the approval of ADC acquisitions. Approvals of ADC acquisitions and AIS programs that include ADC efforts will be based on conformance to the guidelines and standards contained in this ADC Master Plan, compliance with established DoD and Federal life-cycle management (LCM) policies, and certifications by OSD Principal Staff Assistants that the systems are needed.

In addition to demonstrating compliance with DoD and Federal LCM policies, DoD Components will also be required to consider the ADC factors discussed in this Master Plan as they approve, or recommend for approval, acquisitions of systems which include ADC efforts. These factors include specification of ADC-related standards, records management considerations, and development of a business case. (See Sections 2.3, 3.4, and 3.5.)

CONVERSION STRATEGY

OVERVIEW

The Conversion Strategy consists of two components. The first component focuses on the output product of document conversion, *the converted document*. The second, *the management component*, focuses on the management of automated document conversion system acquisitions and the high-level requirements that these systems and services must satisfy.

The Conversion Strategy is intended to be flexible. Conversion output products must be accessible to multiple applications, unless the functional need dictates otherwise. This flexibility permits reuse of converted documents and is achieved by converting documents to approved technical standard formats.

Document conversion may be accomplished by 1) acquiring specialized but commonly available software and hardware, 2) acquiring ADC services from another DoD organization, and 3) acquiring ADC services from the marketplace. The output product of any conversion must conform to the same flexible technical standard format(s) regardless of the acquisition vehicle used.

Document conversion may be viewed as a two-stage process. Each stage produces a conversion output product that conforms to approved technical standard formats. The product of the first stage results in an interoperable baseline(s) and, therefore, is usable across a very wide variety of application software environments. The first stage output is input to the second stage. The product of the second stage conforms to more specialized technical standard formats. These standard formats allow for additional information ("intelligence") about the original document that greatly increases its value as a revisable document. However, this often greatly increases the conversion cost.

The commercial community refers to automated document conversion by the term "imaging." Appendix D provides a primer on imaging.

GOALS

From a technology and acquisition point of view, the Conversion Strategy seeks to:

- maximize the flexibility and utility of converted documents to DoD document users through the requirement to use *de jure* (e.g., ISO, ANSI, IEEE) technical standards
- ensure that the output of a conversion effort supports a DoD mission or business process in a cost-effective manner
- avoid unnecessary duplication in the acquisition of ADC hardware and/or software.

THE CONVERTED DOCUMENT: DEFINITION

A document is a type of record. Historically, the original document medium has typically been paper, microform, or another non-digital form of storage.

Document conversion is the process or activity of moving the information content of a document to a digital format (the converted document) while ensuring preservation of the information. ADC enhances this process by the use of imaging technology to perform the conversion with minimal human intervention.

Digital documents have distinct advantages over analog documents (e.g., paper and microform documents). Digital information may be processed by machine, may be shared through communications links, and in most cases, may be less expensive to locate, retrieve, manipulate, and disseminate.

All ADC systems should apply consistent concepts of document manipulation, including preparation, indexing, scanning, quality control, storage, and retrieval of the converted document throughout the conversion process.

The 'Convert Document' Activity

DoD uses process reengineering to identify military mission and business process improvement opportunities. DoD most often uses the Integrated Definition Language (IDEF) as the methodology for reengineering. The top level in an IDEF process model is the "Context Diagram."

Appendix B describes and illustrates the high-level context diagram for the "CONVERT DOCUMENT" activity and the conversion process. The context diagram for the "CONVERT DOCUMENT" activity reads as follows:

The CONVERT DOCUMENT activity transforms ORIGINAL DOCUMENT into a CONVERTED DOCUMENT. NON-CONVERTED DOCUMENT and SOURCE DOCUMENT are other outputs. This activity uses COST ANALYSIS TOOLS, ADC TOOLS, HUMAN RESOURCES, and FACILITIES (the mechanisms), under the constraints of ADC POLICY and BUDGET.

Two-Stage Document Conversion Process

Digital documents conforming to approved technical standard formats are more easily shared among functional applications. Sharable digital documents are said to be in an interoperable format. Digital documents must not be locked into a proprietary digital format.

In recognition of the cost and long lead times associated with large conversion projects, changing technology, and the potential multiple uses of digital documents, this Master Plan establishes the following two-stage process for document conversion:

- The first stage establishes an interoperable baseline of digitized information. Original documents are converted to a non-proprietary, interoperable format, allowing digital documents to be shared by multiple software applications on different computing platforms.

This first stage conversion process is often labor intensive, because existing paper files must be inventoried, packaged, and scanned. The use of standard, reusable methodologies, procedures, and equipment will minimize these costs. There may need to be more than one interoperable baseline, based on the document types identified in Section 2.3.2. The baseline format may be a raster standard for images, or American Standards Code for Information Interchange (ASCII) for text. (Section 3.4 provides detailed standards guidance.)

- The second stage establishes intelligent document forms that are not satisfied by the interoperable baseline but are required to satisfy mission- or business-justified requirements.

A goal of the second stage is to establish standards for intelligent formats, formats that permit easy document revision. However, a transition period will occur during which proprietary formats may continue in certain applications. The enhanced format may be a text format such as Standard Generalized Markup Language (SGML), or a vector-type standard for engineering drawings.

This flexible two-stage approach extends the potential for reuse of a converted document to satisfy different user requirements for the same document and to develop advanced, intelligent, standards-based converted document formats required to satisfy more demanding mission- and business-justified needs.

First-stage conversion ensures that the information contained in a converted document will be accessible throughout its life, as specified by business needs. The life of the converted document is the same as the original record.

For some applications, it is cost-effective to integrate both stages (first stage and second stage) of conversion into the same conversion procedure. However, such an integration may be counterproductive in the long term. For example, a National Archives and Records Administration (NARA) study examined several conversion programs. The study states that documents converted without a separation between first stage and second stage may be locked into a proprietary format. This situation may cause the converted document to be unusable by other functional applications. Unlocking the information in these documents may require a second conversion, possibly from the original document.

Standardization is a contributor to success of the interoperable baseline approach. For example, the Defense Printing Service (DPS), in conjunction with the Congressionally sponsored Automated Document Conversion System (ADCS) Test, demonstrated the capability to convert baseline documents to application-oriented documents.

Documents that have undergone first-stage conversion become an interoperable baseline from which information can be extracted to meet varying functional requirements. These documents are in a non-proprietary format and are identified through uniform indexing rules. For example, converted paper and microform-based document formats may be prescribed in MIL-R-28002B. Converted medical x-ray formats on the other hand, which could be single-use documents, may require a different standard baseline format. Thus, multiple baselines may exist to satisfy unique functional requirements. It is imperative that any conversion baseline be non-proprietary in format with a uniform index.

Integrating both conversion stages is acceptable if the data format created from the second stage is non-proprietary and interoperable; or if the system for which the data is converted easily converts the data to a non-proprietary and interoperable format.

STANDARDS GUIDANCE FOR CREATING THE CONVERTED DOCUMENT

The DoD is committed to the establishment of a standards-based framework for defining technical architectures. Use of that framework will help ensure interoperability, portability, and scalability of DoD information systems. The framework is identified in the Technical Architecture Framework for Information Management (TAFIM), which includes a set of technical standards. These standards are consistent with recent OSD guidance on maximum utilization of commercial standards.

TAFIM Policy

The TAFIM provides the DoD-wide architecture guidance and a total technical documentation structure for guiding the evolution of the Department's technical infrastructure and future information systems to an open systems environment. It is not a specific architecture for automated document conversion or any other AIS, but rather a framework that establishes the strategic planning direction. It does not set specific business/mission requirements—that is the responsibility of the functional communities. The TAFIM does provide guidelines and statements of direction. It is, therefore, a framework for decision making in an environment of high uncertainty. It provides the services, standards, design concepts, and generic components that will be used to guide architects, designers, and engineers on the selection of compatible configurations leading to more portable, interoperable, and scaleable automated document conversion open systems.

This new strategic direction requires institutionalizing the TAFIM guidance within OSD policy. The ASD(C3I) signed a policy memo dated June 1994 stating that the TAFIM will serve as the single framework to promote the integration of DoD information systems and will guide the evolution of the Department's information system technical architectures. The requirement to use the TAFIM will be incorporated into the updated DoD software management (DoDD 3405.1) and life-cycle management policies (DoD 8120.2-M).

Standards for DoD Automated Document Conversion

All automated document conversion systems must provide output files that conform to the appropriate standards described below, or provide a business case analysis for using proprietary standards and a migration strategy to open systems. These technical standards move the DoD toward open/non-proprietary systems. In a number of instances, additional standards development and testing activities are required in order for the selected standards to fully satisfy DoD systems requirements. See Table I for a summary of the following ADC standards.

1. Image Capture Standards
 - o All digital raster graphics data (black and white bit-mapped images) shall conform to the requirements, scanning resolution, and quality assurance provisions as defined in MIL-R-28002B, 14 December 1992, Requirements for Raster Graphics Representation in Binary Format.
 - o Standards for color and gray scale scanning are still evolving.
2. Image Compression
 - o CCITT Recommendation T.6:1988, Facsimile Coding Schemes and Coding Control Functions for Group 4 Facsimile Apparatus (Group 4) shall be used to encode/decode (compress/decompress) binary raster graphics (black and white bit-mapped images) as defined in FIPS PUB 150 - Telecommunications: Facsimile Coding Schemes and Coding Control Functions for Group 4 Facsimile Apparatus.
 - o JPEG (Joint Photographic Experts Group) (ISO 10918-1) is recommended for encoding/decoding (compression/decompression) of still-frame, continuous-tone, gray scale images that are eight or more bits per pixel, and color images that are 16 or more bits per pixel. JPEG consists of a family of image compression processes. Some data are lost during the processing. (A JPEG "toolkit" allows the user to choose how much loss an application can tolerate.) JPEG provides an extension for true "lossless" compression.
3. Optical Storage Media
 - o CD-ROM standards:
 1. ISO 9660:1988, *Information Processing - Volume and File Structure of CD-ROM for Information Interchange*, covering the logical (data storage) format that makes a Compact Disc readable.
 2. ISO 10149:1989, *Information Processing - Data Interchange on Read-Only 120 mm Optical Data Disks (CD-ROM)*, standardizes the physical characteristics (track shape, track pitch, and data structure) of a Compact Disc.
 - o WORM (Write Once, Read Multiple) standards:
 1. ISO/IEC 9171-1:1990, *Information Technology - 130 mm Optical Disk Cartridge*,

Write Once, for Information Interchange - Part 1: Unrecorded Optical Disk Cartridge.

2. ISO/IEC 9171-2:1990, *Information Technology - 130 mm Optical Disk Cartridge, Write Once for Information Interchange - Part 2: Recording Format*. The standard addresses two formats: Format A, Continuous Composite (CC), and Format B, Samples Servo (SS), are both 325 MB per side but are incompatible with each other. Format A is the ISO/IEC version of ANSI X3.211 while Format B is the ISO/IEC version of ANSI X3.214.
 3. ISO/IEC 10885:1993, *Information Technology - 356 mm Optical Disk Cartridge for Information Interchange - Write Once*. At 3.4 GB per side, this standard is the ISO/IEC version of ANSI X3.200:1992.
 4. ISO/IEC 11560:1992, *Information Technology - Information Interchange on 130 mm Optical Disk Cartridges of the Write Once, Read Multiple (WORM) Type, Using the Magneto-Optical Effect*. At 325 MB per side, this is the ISO/IEC version of ANSI X3.220.
 5. ANSI X3.191:1991, American National Standards Institute - *Recorded Optical Media Unit for Digital Information Interchange - 130 mm Write-Once Sampled Servo RZ Selectable-Pitch Optical Disk Cartridge*. At 650 MB per side, the cartridge dimensions of this standard are different from those of other 130 mm WORM standards.
- o Rewritable Optical Storage standards:
 1. ISO/IEC 10089:1991, *Information Technology - 130 mm Rewritable Optical Disk Cartridge for Information Interchange, Format A - Continuous Composite (CC) and Format B - Sampled Servo (SS)*. Both formats are 325 MB per side, but Format A is incompatible with Format B. Format A is the ISO/IEC version of ANSI X3.212.
 2. ISO/IEC 10090:1992, *Information Technology - 90 mm Optical Disk Cartridges, Rewritable and Read Only, for Data Interchange*. These are 128 MB per side.
4. Data Interchange Standards and Archive Format
 - o Code for Information Interchange, Its Representations, Subsets, and Extensions, (American Standard Code for Information Interchange [ASCII]) (FIPS PUB 1-2) is the standard for the exchange of textual data including any structured arrangement of character-oriented records, files, or indices.
 - o Markup Requirements and Generic Style Specification for Electronic Printed Output and Exchange of Text (Standard Generalized Markup Language [SGML]), which adopts ISO 8879:1986 (MIL-M-28001B for CALS), is the standard for the exchange of textually oriented data.
 - o Requirements for Raster Graphics Representation in Binary Format (Group 4 Raster Scanned Images) (MIL-R-28002B for CALS) is the standard for exchange of data in raster file format.
 - o Digital Representation for Communication of Product Data: IGES (Initial Graphics Exchange Specification) Application Subsets and IGES Application Protocols (MIL-D-28000A for CALS) is the standard for the exchange of 3-D data in vector file format, particularly for documents prepared in CAD/CAM.
 - o Digital Representation for Communication of Illustration Data: CGM (Computer Graphics Metafile) Applications Profile (MIL-D-28003A for CALS) is the standard for exchange of 2-D data in vector file format for documents not prepared in CAD/CAM.
 - o CD-ROM user interfaces, application platform support, and the utility of the information provided continue to be driven by many different producers of CD-ROMs. DoD guidelines for CD-ROM applications are being finalized and will be promulgated throughout DoD.
 5. Marketplace (De Facto) Standards DoD recognizes *de facto* standards (e.g., Tagged Image File Format [TIFF] and Portable Document Format [PDF]) until such time as formal standards are developed and adopted by recognized standards bodies. However, use of a *de facto* standard by an automated document conversion system requires that a business case rationale be developed and a migration plan provided to achieve a non-proprietary, TAFIM-compliant systems environment.

Table I - Summary of ADC Standards

Image Capture	
Optical Storage	
Raster-CALS Raster (MIL-R-28002B)	
<u>CD-ROM</u>	
	ISO 9660
	ISO 10149
Data Interchange	
<u>WORM</u>	
Text-ASCII (FIPS PUB 1-2)	ISO 9171-1
Text-CALS SGML (MIL-M-28001B)	ISO 9171-2
Raster-CALS Raster (MIL-R-28002B)	ISO 10885
2D Vector-CALS CGM (MIL-D-28003A)	ISO 11560
3D Vector-CALS IGES (MIL-D-28000A)	ANSI X3.191
Image Compression	
<u>Rewriteable Disks</u>	
FIPS Pub 150	ISO 10089
JPEG ISO 10918-1	ISO 10090

MANAGEMENT COMPONENT OF CONVERSION STRATEGY

This section provides additional information on how to conduct a business case review based on the conversion strategy. It begins with an analysis of the requirements and an assessment of the mission or business environment to provide the information needed to support the business case. It then addresses issues involved in developing a cost justification or economic analysis. Methods for assessing technical capability of ADC system architectures are also introduced. Finally, this section introduces the key elements of a decision table to assist program managers in determining if ADC for their specific application will meet their operational need and produce sufficient cost savings or cost avoidance to justify the conversion.

Document conversion is a managed activity. The acquisition of equipment, procurement of services, development of technologies, and development of information systems support document conversion. Consequently, such acquisitions must:

- be justified by a business case
- conform to Federal Information Processing (FIP) resource and DoD automated information system management policy

A document conversion program should be considered:

- Only when a military mission or business reason exists for conversion. For collections that include record and non-record material, convert non-record material (non-relevant documents) only if the cost of separating it from record material exceeds the cost of conversion.
- When document conversion is more cost-effective than document re-creation. Convert organization documents of poor quality only if the cost to convert the document contents through Optical Character Recognition (OCR) is lower than direct re-keying. The cost of scanning and correction of OCR errors may be significantly higher than direct re-keying.
- Only when the converted document will be operationally usable in the intended application(s). Conversion programs should focus on documents in which the conversion or end-user system provides adequate image enhancement and editing capabilities to restore any information lost during scanning.

- When the converted document may be accessed by multiple applications. Conformance to standard formats will ensure that multiple applications can access valuable business documents directly or indirectly (i.e., through translators or "bridges").
- When the long-term cost of maintaining or using the document in its original medium exceeds the long-term cost of maintaining the document in a digital medium.
- When the costs of migrating a document over its life include consideration of the long-term stability of the proposed storage medium and the long-term integrity of the storage medium.
- Requirements Determination

An automated document conversion project must be viewed in the context of its mission or business environment. This is accomplished during the development of the business case in support of the requirement.

The requirements (mission need) definition must establish a clear justification for conversion. Cost and performance measures will be developed to evaluate conversion projects.

As a necessary pre-condition to a conversion project, the justification must be approved and meet the following tests:

- *need for information contained in the original documents*

No conversion should be undertaken unless the information contained in the original document(s) is relevant to the supported mission or business function. For example, conversion of technical documents should be limited to original documents that support currently-maintained weapons systems.

- *demonstrated/quantified improvement in the military mission or business process resulting from converting original documents*

This assessment must be made in the context of other considerations, such as the need to protect the original document from permanent loss or the need through document conversion to accelerate an overall business process. For example, the overall cost resulting from a delay in corrective maintenance caused by inadequate access to technical documents may exceed the cost savings realized by not converting technical documents with a low probability of access.

Information Access and/or Dissemination

The choice of indexing schemes and data elements determines accessibility. The use of a uniform indexing scheme, selected to optimize retrieval requirements of converted documents throughout DoD, is essential to the design of consistent interfaces between the conversion system and any end-user systems. Also essential is the identification of local repositories of converted documents. This is the responsibility of the functional official.

The adoption of standard data formats further ensures consistency among converted documents.

Archived Converted Document

The converted document should be stored in an "archival" form to ensure that a converted document, throughout its life, remains available to appropriate applications. Because technology is changing quickly, there is no guarantee that future digital systems may be able to read recordings made on older systems, even if these recordings are still in good condition. In addition, there is no guarantee that *de facto* standards will persist during the life-cycle of either a given system or the converted document produced and stored in that system.

However, conversion projects will need to make provisions for the delivery of digital documents on an appropriate medium, that is in the mainstream of technology. The requirement upon future archives must be limited to only those data and media that are certain to be supported in the out-years. Conversion projects must also provide for the safekeeping of conversion software and hardware in sufficient

quantity to support document access or future reconversion to maintain accessibility, until scheduled document disposition.

The requirements for archiving also need to make provisions for back-up of the converted documents and disposition of the original documents, in accordance with DoD Records Management policies and procedures.

NARA will accept, for permanent records in electronic form, only "one-half inch, seven- or nine-track reel-to-reel magnetic tape and 3480-class tape cartridges" and "CD-ROMs . . . that are in conformance to ISO 9660 standard."

Other Issues

The contents of the converted document must remain available regardless of changes in storage technologies, systems, and applications. NARA recommends reconverting or copying converted records to remain compatible with new storage technologies and to satisfy approved disposition requirements for temporary and permanent records.

Consequently, the sponsor (program manager) must articulate:

- how the converted document will be managed until it has reached the end of the scheduled retention period (conform to NARA policy)
- how the intended application may impact future reconversion requirements
- what migration path, such as the adoption of integrated processes and interactive documentation, will be used to accommodate changes in information systems
- Cost Justification

Requirements for cost justification will include operational, financial, and long-term concerns:

- **Operational.** The operational justification must state that the converted document will support the mission or business requirement and that the information contained in the document cannot be obtained from another source in a more cost-effective manner.
- **Financial.** The financial justification must state the financial impact of the conversion process. Additionally, the justification must indicate whether to contract for conversion services, to buy conversion hardware and software, or to use a centralized DoD conversion service.
- **Long-Term.** The long-term costs associated with managing media for storing electronic documents, hardware to access the media, software to read the media, and migration to newer media due to deterioration of old media must be well understood. Planning should include these costs for the life-cycle of the information being managed on electronic media.
- Technical Capability

The technical capability of an ADC system will be determined by an evaluation of the system's architecture and its conformance to published standards. (A discussion of the TAFIM and ADC standards guidance is provided in Section 3.4.)

Business Case Decision Table

The decision table below is provided as a guide to assist program managers in determining if the proposed ADC acquisition will:

- meet operational requirements
- produce sufficient cost savings or avoidance to justify the conversion

Table II - Business Case Decision Table Requirements Determination

- Is there a legitimate mission or business need?

- Will Records Management requirements be satisfied?
- Are the records scheduled per NARA requirements?
- Are NARA archival requirements satisfied?

Cost Justification

- Can the information contained in the documents be obtained from another source in a cost-effective manner?
- Will automated document conversion reduce costs?
- Has a comparison been made of purchase vs. contracting the automated document conversion service?
- Should centralized DoD conversion services be considered?

Document Candidate Selection

- Are the documents active and do they have sufficient volume?
- Are the documents available to multiple users?
- Do the documents contain valuable and relevant information?
- Do the documents have a relatively long active life remaining?
- Are the input/information processing routines stable?
- Can the original documents be destroyed after conversion?

Technical Capability

- Architecture
 - Does the selected architecture support the minimum functionality, ensuring the interoperability of the converted documents?
- Standards
 - Are relevant standards identified to ensure interoperability of the converted documents?

Once the decision is made to proceed with automated document conversion, the following should be used as conversion strategy guidance:

- First Stage Conversion
 1. Use the DoD standards from Table I or provide a business case for use of proprietary standards and a migration strategy to open systems.
 2. Evaluate standards to ensure they meet the interoperability requirements and use non-proprietary formats.

If first-stage document conversion meets the end user's functional/mission requirements, proceed with LCM documentation. If the conversion does not meet requirements, then move to second stage conversion.

- Second Stage Conversion
 1. Tailor document conversion to meet end-user requirements.
 2. Evaluate standards to meet interoperability requirements and use non-proprietary formats.

If interoperability cannot be achieved with second stage conversion, evaluate the cost of maintaining an interoperable baseline copy after first stage conversion.

ROLES AND RESPONSIBILITIES

This section assigns roles and responsibilities for DoD officials involved with ADC. The assignment is based on the management concept of centralized policy development and decentralized execution.

Assistant Secretary of Defense (C3I)

The ASD(C3I) is the DoD Senior Information Management (IM) Official. The ASD(C3I) leads and establishes information management policy for the Department. The responsibility of the ASD(C3I) is to:

- develop DoD document conversion policy
- oversee the acquisition of ADC programs designated as major AISs
- review and approve (or delegate authority to approve) ADC acquisitions that use FY 1995 funds and exceed \$5,000,000.

In support of these responsibilities, ASD(C3I) will develop an action plan that addresses follow-on activities such as finalizing standards and providing additional implementation guidelines. The Defense Information Systems Agency (DISA), as the technical support organization for ASD(C3I), will lead the activity to define future ADC standards and guidance.

Functional Officials

The senior functional officials of the Department are the OSD Principal Staff Assistants (PSAs) to the Secretary of Defense. The OSD PSAs establish policy for the Department in their respective functional areas of responsibility. The document conversion role of a functional official is to:

- implement DoD policy to maintain, protect, and preserve organization records
- determine the business need for document conversion
- support efforts to improve the accessibility to functional documents
- validate the business need and justification for document conversion

Development of automated document conversion functional process and data models is the responsibility of DoD functional officials in accordance with CIM guidelines.

DoD Components

The DoD Components include the Military Departments, the Defense Agencies and field activities, the Joint Staff, and Unified and Specified Commands. The document conversion role of a DoD Component is to:

- implement DoD policy to maintain, protect, and preserve organization records
- determine the Component-unique mission and business needs for document conversion
- support efforts to improve the accessibility of documents
- validate the Component-unique mission and business need, and justification for document conversion
- convert documents internally, use centralized DoD conversion services, or contract out for conversion services, ensuring employment of the most cost-effective conversion approach consistent with the business justification
- Commanders, Managers, Records Officers, and Document Custodians

These officials must:

- coordinate with the functional manager, records officer, systems support officer, and legal officer to determine eligibility of an original document for conversion
- determine if the original or the converted document will be the organization record, in accordance with the organization's published records disposition schedules approved by the Archivist of the United States.
- consider the long-term financial costs of managing documents in either their original medium or in their converted medium
- consider the stability and life-cycle of the proposed medium of storage for converted documents
- consider the life-cycle of the hardware and software used to store, retrieve, and manage a

- converted document throughout its life
 - select documents for conversion based on the information value, need to share the information in an electronic environment, and number of times the information will be accessed over its life
-

DoD ADC Master Plan

Appendices

APPENDIX A - EXAMPLES OF BUSINESS REQUIREMENTS FOR AUTOMATED DOCUMENT CONVERSION

This appendix discusses functional requirements and current automated document conversion systems, and planning issues associated with document conversion projects.

1. BUSINESS REQUIREMENTS FOR DOCUMENT CONVERSION

Like other organizations, DoD has a basic requirement for document conversion. This section cites common private sector reasons for conversion, and provides examples of DoD conversion efforts.

1.1 PRIVATE-SECTOR BUSINESS REQUIREMENTS

The most important reasons given in the industry survey cited in Section 2.2 for converting business documents are:

- "better movement and access of documents"
- "improved productivity"
- "reduced employee head count"
- "better management control over process"
- "competitive advantage"
- "improved security of documents"
- "reduced storage space costs"

These reasons provide a business context and justification for conversion. The driver of digitization is the desire to better manage information and business processes. Efficiency and cost savings are the business goals. This is consistent with DoD Corporate Information Management (CIM)/Enterprise Integration goals:

- "reinvent" and reengineer DoD functional processes to achieve greater mission effectiveness at lower costs
- tie DoD together through the use of common, shared data
- minimize duplication and enhance DoD's information systems to embody reengineered processes

An imaging and conversion industry trade association recognizes that document conversion/imaging is an enabling step toward improving business performance.

"Paper creates the biggest bottleneck in the operation of most businesses, governments, and institutions today. Checks, correspondence, invoices, purchase orders, engineering drawings, and forms of all types are, for the most part, processed by hand."

"This paperwork problem is actually part of a larger business problem—the requirement to develop methods to increase the productivity of white-collar staff (e.g., engineers, bankers, secretaries, lawyers, and managers) and to improve the quality of their output. To achieve these goals, the entire business process of an organization must be reviewed and usually changed."

Viewing the references to paper and paperwork as broadly applicable to all analog forms of documents, the comments above apply equally to the concerns addressed in this Master Plan.

1.2 DOD BUSINESS REQUIREMENTS

These same business drivers and goals are motivating DoD document conversion efforts. Key examples from documentation of existing and planned projects are discussed below.

1.2.1 Personnel

A comprehensive set of requirements that includes improved information management, improved business functionality, and integration of business information across systems is illustrated in the following personnel example:

"The purpose of DPRIS is to . . . provide for timely and accurate update of the member record; to provide access to these records for authorized users in a timely and accurate fashion; to protect these records from unauthorized use and inadvertent disclosure . . . ; to obtain the greatest possible integration of the records system with other personnel accounting system[s]."

1.2.2 Finance

Using document imaging as the supporting framework for improving business performance is the approach chosen by the financial community.

"The [Defense Finance and Accounting Service (DFAS)] Imaging Program is designed to improve the management of customer services and reduce the overall operational cost of DoD financial activities."

"The DFAS Imaging Program will support the finance and accounting network consolidation . . . and will facilitate work load transfers by converting large paper-based information stores to digital document repositories and databases." Program objectives include:

"Invest in imaging and scanning capability to reduce the volume of paper required in the business centers and improve management of work flows through automated gathering and routing systems."

"Reduce the cost, manpower and space requirements for the near and long term retention of required backup data with the deployment of high capacity magnetic, optical and video storage media."

"Introduce indexing and searching of stored bit mapped or digital image data to improve customer service levels . . . "

"Improve customer service through innovations in document image sharing and on-line matching and viewing of images with data systems."

"Reduce or eliminate expenses for photo copies, manual sorting and routing, fees and penalties for late payments, and duplicative data entry and document retention."

1.2.3 Engineering

Management, control, and dissemination of information are the driving requirements in the engineering community.

"[A] significant increase in demand for engineering drawings and associated data" dictated "that measures be taken to effectively manage and control engineering drawings and associated data in a cost effective manner." "JEDMICS [Joint Engineering Data Management Information and Control System] provides engineering data and related repository services to support DoD activities which require such data." Bulk conversion of aperture cards, as well as hardcopy and film, to a digital data format is required.

2. CONVERSION OBJECTIVES

Current projects that involve document conversion tend to have the common goal of implementing automated document management systems to reduce the need to print, copy, distribute, and store paper. Such systems will also provide rapid electronic access to data, and electronic sharing, monitoring, and storage of documents. Several types of common objectives were observed.

2.1 COMMON, INTEROPERABLE SOLUTIONS

Automated document conversion planning is being done as part of the DoD's movement toward achieving CIM and Enterprise Integration goals. Common, interoperable solutions are being targeted.

Finance. "Aggressively apply new methods and technologies . . . The program meets the Defense Management Review Decision (DMRD) 918 requirement to migrate toward open systems architecture, and accelerate consolidation and standardization efforts for business process improvements."

2.2 ENTERPRISE AND DESKTOP INTEGRATION

There is a strong business demand for "desktop application integration" and "enterprise document management." These are two key areas where the Gartner Group reports that even vendors are pushing for standards, as organizations shift from a paper-based office "to rely increasingly on digital access, retrieval, viewing and storage of documents."

Personnel. Move "closer to a more common, joint arena. . . [P]ursue a migration strategy toward a common imaging system." "Also under way is the data and forms standardization which will permit the conversion of personnel records into standardized electronically processed forms. As this conversion to paperless forms is completed, the optical imaging of paper forms will be eliminated."

2.3 DOCUMENT INTEGRITY AND REUSABILITY

Another major industry issue, ensuring "the integrity and reusability of electronic document assets," is also a major driving force behind DoD conversion programs.

Engineering. JEDMICS "will use digital techniques to store, retrieve, reproduce, and distribute engineering drawings presently retained in the form of aperture cards, microfilm and hard copy." It will automate "repositories and technical libraries with an optical disk-based digital storage and retrieval system using off-the-shelf hardware and software. Images will be retrieved from the system with the aid of contractor provided application software, and distributed on a variety of media depending on the needs of the ordering activity."

3. SUMMARY

Many automated document conversion systems are installed or near completion. While these "legacy" systems have been designed to meet certain functional requirements, they were conceived prior to the existence of this Master Plan. As DoD moves toward the goal of interoperable baselines, existing systems must be evaluated for duplication in capability, incompatibility in output product formats, and conformance to standards.

DoD strategies and planning efforts for the future should continue to assess candidate conversion/imaging programs from a business perspective, in which "document imaging should be viewed as an enabling suite of technologies rather than as a solution in and of itself."

APPENDIX B - CONVERT DOCUMENT PROCESS MODEL

1. BUSINESS PROCESS REENGINEERING

DoD uses process reengineering to identify business process improvement opportunities. DoD most often uses the Integrated Definition Language (IDEF) as the methodology for reengineering. The top level in an IDEF process model is the "Context Diagram" (Figure B-1). The context diagram describes the most general activity (only one activity is named at the top level of the model), inputs, controls, outputs, and mechanisms relevant to the purpose of the model.

2. CONVERT DOCUMENT

An activity transforms inputs into outputs and is performed by mechanisms under the constraints set by controls. The context diagram for the "CONVERT DOCUMENT" activity reads as follows:

The CONVERT DOCUMENT activity transforms ORIGINAL DOCUMENT into a CONVERTED DOCUMENT. NON-CONVERTED DOCUMENT and SOURCE DOCUMENT are other outputs. This activity uses COST ANALYSIS TOOLS, ADC TOOLS, HUMAN RESOURCES, and FACILITIES (the mechanisms), under the constraints of ADC POLICY and BUDGET.

Figure B-1 is the high level context diagram for the "CONVERT DOCUMENT" activity.

3. DOCUMENT CONVERSION MODEL

Figure B-2 depicts the change in status of the ORIGINAL DOCUMENT at the end of the conversion process. The model shows that the ORIGINAL DOCUMENT is screened and selected for possible migration to a medium other than its original. The ORIGINAL DOCUMENT at the end of the conversion process produces four potential outputs. The first (NON-CONVERTED DOCUMENT) is reviewed but not selected for conversion. The second (NON-MIGRATABLE DOCUMENT) has been selected for conversion but due to technological limitations cannot be converted (e.g., size, shape, quality). Two outputs, NON-CONVERTED DOCUMENT and NON-MIGRATABLE DOCUMENT, exit the process unchanged. The third (CONVERTED DOCUMENT) is the electronic or digitized version of the ORIGINAL DOCUMENT. The fourth (SOURCE DOCUMENT) is the ORIGINAL DOCUMENT physically unchanged that has passed through conversion process and was used to create the CONVERTED DOCUMENT.

FIGURE B-1 - CONVERSION CONTEXT DIAGRAM

FIGURE B-2 - CONVERSION PROCESS

APPENDIX C - DEFINITIONS

ADC Policy	ADC guidelines that provide criteria for migrating a document from one medium to another medium.
ADC Tools	Hardware, software, and other devices used to migrate a document from one medium to another medium.
Administrative Record	Records relating to budget, personnel supply and similar housekeeping or facilitative functions common to most agencies.
Budget	The funds programmed for document conversion.
Convert Document	The process of migrating a document from its current medium to another.

Converted Data	A second-generation copy of data in digital form.
Converted Document	A document migrated from one medium to another. The second-generation copy of an original document having been converted to digital form (digitized document).
Cost Analysis Tools	Methods (automated or manual) used to evaluate cost.
Data	Information used as a basis for reasoning, discussion, calculation, and decision making.
Digital Data	A copy of data in a binary or numerical form that is interpreted by a computer. Often referred to as "electronic" data.
Document	A type of record. Original form historically paper or other non-electronic medium.
Document Conversion	The process of migrating a document from its original medium of storage to another medium of storage.
Document Management	The life-cycle management of documents in accordance with approved records management policies, processes, and procedures.
Facilities	Where a document is managed.
Formats	Specific characteristics (shape, size, style), organization, arrangement, or general layout of a collection of information.
Human Resources	People.
Imaging	The technology that uses scanners to capture and convert documents to a digital form.
Information	Facts or knowledge communicated or received.
Intelligent Document	A document (in digital form) containing, in addition to its primary data, embedded data used to enhance use of the primary data and/or facilitate conveyance of the document's information.
Medium	The environment on which information resides (e.g., microform, electronic, paper, etc.).

Non-Converted Document A document NOT selected for migration.

Non-Migratable Document A document not converted due to technological limitations.

Organization Record Records related to the tasks, functions, mission, or operation of an office. These are divided into two subgroups, administrative and program records.

Original Document A document being managed as the organization record.
The first generation version of a document from which copies or reproductions are made.

Permanent Record Records determined by NARA to have sufficient historical or other value to warrant preservation beyond the time they are needed for administrative, legal, or fiscal purposes.

Program Record Records documenting the unique substantive functions for which an agency is responsible.

Record A record consists of information, regardless of medium, detailing the transaction of business.

Records Management The planning, controlling, directing, organizing, training, promoting, and other managerial activities involving the life-cycle of information, including creation, maintenance (use, storage, retrieval) and disposal, regardless of media.

Screen a Document Review for compliance with conversion policy.

Source Document The original document that has passed through the conversion process.

Work Cards Documents listing detailed maintenance procedures used by personnel in preventive maintenance activities. Synonymous with maintenance record cards.

Workflow The movement of work (forms or documents) through an organization.

APPENDIX D - BASIC ELEMENTS OF ELECTRONIC DOCUMENT CONVERSION

This appendix assists the decision maker by providing an explanation of some of the basic elements of imaging and document conversion.

1. THE BASIC PROCESS

Document image processing refers to the capture, storage, and retrieval of information in the form of electronic images. The basic imaging process is illustrated in Figure D-1.

FIGURE D-1 - THE BASIC IMAGING PROCESS

It begins with document preparation and *scanning*, which converts the hard copy image into an electronic image file. Specialized image processing hardware or software then compresses the image file. *Image compression* reduces the size of an image file by eliminating unnecessary or redundant data. The image file is then *indexed*, either manually or using *optical character recognition* (OCR). OCR is a technology that translates printed characters into machine-readable text. The index is stored in a database and is used to retrieve the image for later viewing. A *quality control check* is performed by an operator and ensures that the image has been properly scanned and indexed. The images are usually stored on *optical disks*. Optical disks are high-density storage devices that are written to and read by laser light. Images are retrieved using the index database and can be viewed on monitors or printed.

1.1. RASTER IMAGE FILES

The image file is different from a standard text file, because it stores information in the form of *raster* images rather than ASCII data. The basic unit of information in a digital computer is the *bit*, which represents a single "on" or "off" switch. In traditional data processing, the ASCII standard is used to translate patterns of bits (arranged into 8-bit *bytes*) into letters, numerals, and a variety of special characters. These letters and numerals can then be discretely read or otherwise manipulated (delete, add, multiply, etc.) by computer software. In imaging, patterns of on or off bits are interpreted as *pixels*, or picture elements, which are analogous to one dot on a display screen. In the simplest case, a pixel would be interpreted as either black or white, and be represented by a single on or off bit. Computers use imaging software and standards to translate bit streams into *arrays* (columns and rows) of pixels that form pictorial representations, or *images*, of textual, numeric, and other information. (See Figure D-2).

FIGURE D-2 - ASCII TEXT VS. IMAGING

Although the distinction between ASCII text and image files may seem purely technical it has a profound impact on information accessibility, manipulation, and security. The ramifications of images versus data processing include:

Image files capture pictorial as well as textual data. This enables them to store signatures as well as graphics, photographs, and drawings.

Pixel elements have no intrinsic information value and, therefore, cannot be interpreted directly by a computer. The translation of image representations into useful information is performed in the mind of the user.

The chief means of input to electronic document management systems is scanning, as opposed to data entry. This provides unique challenges as well as opportunities.

The use of WORM (Write Once Read Many) technology in electronic document management provides a new dimension of information security.

Image files can be viewed as another type of data. However, the large sizes of image files create new processing challenges.

1.2. VECTOR IMAGE FILES

Vector images are an alternative means of representing graphical information on a computer. They are used in CAD (Computer-Aided Design) systems to store engineering drawings. Vector images are stored as a series of descriptions of straight lines defined by their end points. Although they cannot be read directly by electronic document management systems, CAD files can be converted to raster images. Alternatively, some image files can be converted to CAD files in a process called *vectorization*. However, it should be noted that the vectorization software is still quite expensive, and the conversion requires significant manual effort as well.

1.3. IMAGE QUALITY

One measure of image quality is resolution in *dots per inch* or DPI. Dots per inch refers to the number of pixels used to define an image. One-hundred DPI would mean 100 x 100 or 10,000 dots per square inch of image. Resolutions of 200 to 400 DPI are sufficient for most business applications. At 200 DPI an 8.5" by 11" document will require 8.5 x 11 x 200 x 200 or 3,740,000 pixels to define its image.

The number of bits required to represent a single pixel is dependent on the amount of *gray scale* information required by the system. Gray scale refers to the shades of black that can be represented. As shown in our previous example, the simplest case is either black or white and requires only a single bit per pixel. An image composed of black and white dots is called a *halftone* image. Most scanners are designed to detect between 4 and 256 shades of gray. The storage of 4 shades of gray would require 2 bits to represent a single pixel. To store 256 shades would require 8 bits per pixel. Because the use of gray shades can dramatically increase file sizes, gray scale requirements should be kept to a minimum. *Dithering* is a technique that simulates gray tones in a halftone image through altering the size, arrangement, or shape of dots. Unless a document requires gray scale (e.g., black and white photographs), dithering may be a cost-effective alternative.

In an uncompressed format, a 8.5" x 11" document at 200 DPI with 8 levels of gray will require 11,220,000 (8.5 x 11 x 200 x 200 x 3) bits or 1,402,500 (11,220,000 / 8) bytes (1.403 MB) to store its image. In order to reduce storage and transmission costs *image compression* is used to reduce the amount of storage necessary for each document. Image compression is performed by specialized hardware or software and uses a variety of algorithms to remove bits that define blank space or redundant data. Compression ratios of 1:20 are not uncommon. With that kind of performance our example's storage requirement would be reduced from 1.403 MB to 70K.

1.4. STORAGE DEVICES

The proliferation of electronic document management is in large part due to the development of high-density, inexpensive optical storage devices. The three major types of optical media are *CD-ROM* (Compact Disc-Read Only Memory), *WORM* (Write Once Read Many), and *rewritable optical* (also known as magneto-optical). Storage capacities from hundreds of megabytes to gigabytes are available on single disks. The features of these devices are described in the following section along with other technologies involved in electronic document management.

2. SCANNING TECHNOLOGY

The technologies used in electronic document management include scanners, storage devices, printers, monitors, control computers, software, and networks.

2.1. SCANNERS

Scanners are used to translate printed images into digital machine-readable files. Light from a high-intensity lamp is directed at the target document. The light reflected from the document is directed by a set of mirrors and lenses to a photodetector. The photodetectors are usually arrays of charged-couple devices (CCDs). CCDs use semiconductor chips with light-sensitive elements to measure the intensity of the reflected light and convert it to an analog electrical signal. The scanner's

controller translates the analog signal into a digital pattern of bits that can be passed to the control computer.

The first desktop scanners could record only halftone images; but today most scanners will detect 256 or more gray tones. Color scanners use color filters or lamps to record color images. There are also specialized scanners that can scan microfilm, microfiche, aperture cards, and large format (E-size) drawings. A further innovation in scanning is the duplex scanner. A duplex scanner can scan both sides of a document at one time as opposed to single-side, or simplex, scanner.

Scanners use various mechanisms to guide documents through the system. The flatbed scanners are similar to conventional photocopy machines. The document lies on a glass plate and a system of mirrors guides the light to the document and directs the reflection to the CCD array.

In sheet-fed scanners (also called moving-paper or edge-feeding scanner) the source document is moved over the scanning device. Many vendors have also introduced automatic document feeders.

Scanning software is required to interpret the bit stream produced by the scanner controller. This software can be resident in the scanner itself or on an external control computer. The extent and sophistication of scanning software included with the scanner varies from product to product.

2.2. STORAGE DEVICES

Because of their high information storage densities, optical disks are the storage media of choice for electronic document management systems. The three basic types of optical storage devices are CD-ROM (Compact Disk-Read Only Memory), WORM (Write Once Read Many), and rewritable optical. WORM and rewritable optical are the technologies used by electronic document management. CD-ROM is used mainly in publishing.

3. IMAGE CAPTURE

Document capture consists of the conversion of documents to enable their use in electronic document management systems. Subtasks include document preparation, scanning, and indexing. A generic workflow chart for document capture is illustrated in Figure D-3.

FIGURE D-3 - DOCUMENT CAPTURE WORKFLOW

4. STORAGE

The storage process includes the compression of image files, the writing of the files to optical storage, and the updating of the index database. A chart illustrating the workflow of the storage process is contained in Figure D-4.

FIGURE D-4 - IMAGE STORAGE WORKFLOW

APPENDIX E - CONVERSION GUIDES FOR PROGRAM MANAGERS

This series of guides is designed to provide managers with an understanding of the technical issues that must be addressed through a document conversion life-cycle. Because immediate conversion priorities are for paper/microform conversion to digital documents, the guides focus on this type of conversion. The guides, which will be published separately, cover the following topics:

- **Conversion Project Components**

For the purpose of the ADC Master Plan, conversion can be broken into five components, namely justification, operation, management, quality control, and audit. This technical guide analyzes these components in the context of a paper/microform conversion. In particular, this guide shows the place that each of these components has in the life-cycle of the converted documents. Of special attention is

the different importance that each of these components has in a first or second stage conversion.

- **Conversion Hardware and Software**

This guide deals with the tools that support conversion project components, in the context of the automated document conversion system. To that end, it addresses mainly the items that are necessary for document-centric (first stage) conversion for an interoperable baseline. It also reviews some of the application-centric (second stage) conversion components that will make use of the converted documents to meet a specific application requirement. This guide also provides trade-off analysis between hardware and software technical solutions, so that managers can optimize the resources they need to deploy in a given conversion operational environment.

- **Conversion Workflow**

This guide will link the operation, quality control, and audit activities described in the *Conversion Project Components* guide, as they relate to a conversion project. This guide will be in two parts, a conversion project workflow, dealing with the actual conduct of a conversion project, and an operational support workflow, related to the maintenance and management of a conversion installation.

The conversion project workflow will describe in detail the steps that occur during a conversion (document receipt, inventory, planning, triage, preparation, indexing, scanning, quality control, error corrections, storage, documentation, packing, release), and the various possible sequencing of these steps to reflect different operational and document use constraints.

The operational support workflow will cover the activities necessary for a smooth operation, regardless of the nature of the conversion projects being conducted. These activities will be essentially of an audit nature, dealing with operations (such as supply procurement, periodic maintenance, and compliance with policies), project management, and technical controls.

- **Conversion Metrics and Standards**

This guide will provide the information necessary for a manager to ascertain the soundness of a particular conversion project, and the operational pointers that will indicate whether a project is progressing according to expected performance levels.

It will provide industry-accepted performance and technical metrics for different types of conversion systems, such as specifications for various equipment, as well as cost factors for different operations.

APPENDIX F - ACRONYMS

ADC	Automated Document Conversion
ADCS	Automated Document Conversion System
AIS	Automated Information System
ANSI	American National Standards Institute
ASCII	American Standard Code for Information Interchange
ASD (C3I)	Assistant Secretary of Defense, Command, Control, Communications, and Intelligence
CALS	Continuous Acquisition and Life-Cycle Support

CCITT	Consultative Committee on International Telegraph and Telephone
CD-ROM	Compact Disc - Read Only Memory
CFR	Code of Federal Regulations
CGM	Computer Graphics Metafile
CIM	Corporate Information Management
DISA	Defense Information System Agency
DPS	Defense Printing Service
FIPS	Federal Information Processing Standard
IDEF	Integrated Definition Language
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IGES	Initial Graphics Exchange Specification
ISO	International Standards Organization
JEDMICS	Joint Engineering Data Management and Information Control System
JPEG	Joint Photographic Experts Group
LCM	Life-Cycle Management
MB	Megabyte
NARA	National Archives and Records Administration
OCR	Optical Character Recognition
PDF	Portable Document Format
PSA	Principal Staff Assistant
PUB	Publication
SGML	Standard Generalized Markup Language
TAFIM	Technical Architecture Framework for Information Management
TIFF	Tagged Information File Format
WORM	Write Once, Read Many

APPENDIX G - REFERENCES

- 36 Code of Federal Regulations.
- 44 United States Code.
- ANSI/AIIM TR-2, 1992.
- ASD(C3I) Memorandum, *Technical Architecture Framework for Information Management (TAFIM)*, 23 June 1994.
- Avedon, Don M., *Introduction to Electronic Imaging, Second Edition*, Association for Information and Image Management, 1994.
- *Department of Defense Records Management Task Force Report One*, January 1995.
- Department of Defense Records Management Task Force Report, *Refinement of the Generic Entity "RECORD,"* March 1995.
- DISA Center for Standards, *Proposed Imagery Standardization Management Committee (ISMC) Charter*, 2 December 1994.
- DISA, *Technical Architecture Framework for Information Management (TAFIM)*, Volumes 1-3, October/December 1992.
- DoD Directive 3405.1, *Software Management*, (Draft for Comment), 4 November 1994.
- DoD Directive 5015.2, *Records Management Program (Change 1)*, 3 April 1992.
- DoD Directive 8000.1, *Defense Information Management (IM) Program*, 27 October 1992.
- DoD Directive 8120.1, *Life-Cycle Management (LCM) of Automated Information Systems (AISs)*, 14 January 1993.
- DoD Instruction 8120.2, *Automated Information System (AIS) Life-Cycle Management (LCM) Process, Review, and Milestone Approval Procedures*, 14 January 1993.
- DoD 8120.2-M (Draft), *Automated Information System Life-Cycle Management Manual*, 3 October 1994.
- DoD, *Corporate Information Management for the 21st Century - A DoD Strategic Plan*, April 1994.
- DoD, *Defense Information Infrastructure Master Plan*, Version 1.0, 22 November 1994.
- DoD, *Information Technology Standards Guidance (ITSG)*, Version 2.0, 20 September 1994.
- DPS/DLA/DISA, *Test Analysis Report of Automated Document Conversion Tests and Recommended Applications of Automated Document Conversion Systems*, 12 December 1994.
- FIPS 150, *Facsimile Coding Schemes and Coding Control Functions for Group 4 Facsimile Apparatus*, 4 November 1988.
- FIPS 183, *Integration Definition for Function Modeling (IDEF0)*, 21 December 1993.
- FIPS 184, *Integration Definition for Information Modeling (IDEF1X)*, 21 December 1993.
- Kirwan, Mary Jo and Laurel Worm, *Electronic Document Management System Guidelines*, Universal Systems, Inc.
- MIL-R-28002B, *Raster Graphics Representation in Binary Format, Requirements for*, 14 December 1992.
- NARA Bulletin 94-4, 20 July 1994.
- NARA Technical Information Paper No. 12, *Long Term Access Strategies for Federal Agencies*, July 1994.
- NARA, *A Federal Records Management Glossary*, Second Edition, 1993.
- National Defense Authorization Act for Fiscal Year 1994, House Report 103-200.
- NML Technical Report RE-0016, *Storage Technology Assessment Final Report*, July 1994.
- NML Technical Report RE-0017, *Media Stability Studies Final Report*, July 1994.
- Popkin, J., *Integrated Document Management: Key Issues for 1995*, Gartner Group Research Note, 30 November 1994.
- Public Law 103-335, Section 8114, *Defense Appropriations Act of 1995*, 30 September 1994.
- Secretary of Defense Memorandum, *Specifications & Standards-A New Way of Doing Business*, 29 June 1994.
- The Applied Technologies Group, *The Complete Guide to Document Imaging*, Document Imaging Executive Information Report Series, Wang Certified Image Consultant Program, 1993.